Ana Salgado/Rute Costa/Toma Tasovac

APPLYING TERMINOLOGICAL METHODS TO LEXICOGRAPHIC WORK: TERMS AND THEIR DOMAINS

Abstract Applying terminological methods to lexicography helps lexicographers deal with the terms occurring in general language dictionaries, especially when it comes to writing the definitions of concepts belonging to special fields. In the context of the lexicographic work of the *Dicionário da Língua Portuguesa*, an updated digital version of the last Academia das Ciências de Lisboa' dictionary published in 2001, we have assumed that terminology – in its dual dimension, both linguistic and conceptual – and lexicography are complementary in their methodological approaches. Both disciplines deal with lexical items, which can be lexical units or terms. In this paper, we apply terminological methods to improve the treatment of terms in general language dictionaries and to write definitions as a form of achieving more precision and accuracy, and also to specify the domains to which they belong. Additionally, we highlight the consistent modelling of lexicographic components, namely the hierarchy of domain labels, as they are term identification markers instead of a flat list of domains. The need to create and make available structured, organised and interoperable lexicographic resources has led us to follow a path in which the application of standards and best practices of treating and representing specialised lexicographic content are fundamental requirements.

Keywords Definition; domain label; general language dictionary; lexicography; term; terminology

1. Introduction

The title of this paper highlights our belief that terminology as a science with its own methodology and interdisciplinary and transdisciplinary nature (Felber 1987, p. 1) can contribute to a practice-based rethinking of lexicographic work when terms are at the core of the analysis. We will demonstrate on these pages that terminological methods can help lexicographers and are advantageous for the process of lexicographic knowledge-building.

Due to the democratisation of knowledge, the growth of communication media and the technological and scientific boom, terms are exceptional sources of lexical renewal and enrichment of the language systems. Thus, their registration in general language dictionaries has increased over the years (Rondeau 1984, pp. 1–4).

Many researchers have conducted studies on the presence of terms in general language dictionaries based on monolingual dictionaries (Rey 1985; Béjoint 1988; Tournier 1992; Cabré 1994; Paz Battaner 1996; Estopà 1998; Boulanger 2001; Roberts 2004; Guerra Salas/Gómez Sánchez 2005; Nomdedeu Rull 2008), reviewing different topics (e. g., coverage and percentage of terms, domain labelling, terms related to a specific domain, etc.). We distance ourselves from these authors whenever we apply terminological methods to lexicographic work since we believe that lexical units (words in general) and terminological units (terms) must be differentiated. Lexicography and terminology are two disciplines with different theoretical and methodological assumptions and whose final products aim to respond to different social needs. In this context, we will describe the method we apply to treat terms in general language dictionaries, mainly based on International Organisation for Standardisation (ISO) standards, namely ISO 704 (2009) and ISO 1087 (2019).

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

In the universe of the labelling system commonly used in lexicography, labels assigned to specialised senses are called domain labels, a 'marker which identifies the specialised field of knowledge in which a lexical unit is mainly used' (Salgado/Costa/Tasovac 2019). These markers represent the most efficient method to detect terms in general language dictionaries, which justifies our interest in this type of label.

For the sake of consistency, throughout this paper, we have adopted some typographic conventions as exemplified below:

- Domain labels are written in small caps, e. g., GEOLOGY.
- Terms are written in quotation marks, e. g., "term".
- Concepts are written in angled brackets and with the first letter capitalised in a fixed-width (monospace) font, e. g., `<Concept>`.
- Concept relation identifiers are written with an underscore between the forms in a fixed-width (monospace) font, e. g., `is_a`.
- TEI P5[1] terms are written in a fixed-width (monospace) font.

The rest of this paper is organised as follows. Section 2 aims to start by clarifying some of the key concepts, namely term, which necessarily brings concept along, and, subsequently, we conduct our research in light of the double dimension of terminology. Section 3 presents our dictionary case study and the domain selected for the study (GEOLOGY). Section 4 is dedicated to the applied terminological working methods used in a Portuguese language dictionary. Finally, we present concluding remarks and highlight our future work.

2. Framework issues

Term and concept are two core keywords that have been defined quite differently by the various theoretical approaches in terminology (e. g., Wüster 1979/1998; Felber 1987; Cabré 1999; Temmerman 2000; Gaudin 2007; Faber 2009). Despite that, we adopted ISO definitions, i. e., a *term* is understood as a '*designation* that represents a *general concept* by linguistic means' (ISO 1087 2019, p. 7), and a *concept* 'should be viewed not only as a unit of thought but also as a unit of knowledge' (ISO 704 2009, p. 3) 'created by a unique combination of *characteristics*' (ISO 1087 2019, p. 3). In other words, the *concept* – a non-linguistic element – is designated by the term, and the *term* – a linguistic element – in turn lexically designates the concept (Fig. 1).

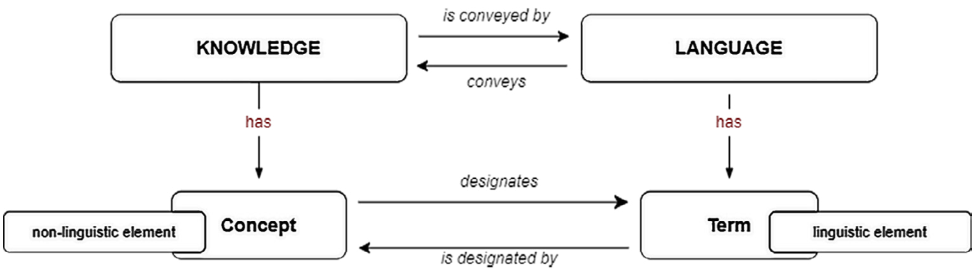

Fig. 1: The Relationship of Concept and Term mirroring the double dimension of terminology (adapted from Costa 2021)

[1] https://tei-c.org/guidelines/p5/.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

We always bear in mind that terms lexically designate concepts, which are often not of primary concern to lexicographers, who usually start from the word form to identify senses, pushing the concept to a secondary level, or ultimately disregarding it. Instead of following this semasiological approach, we propose a different and combined perspective prioritising the concept.

Since the concept 'is created by a unique combination of *characteristics*' (ISO 1087 2019, p. 3) we need to know that a *characteristic* is an 'abstraction of a *property*' (ISO 1087 2019, p. 3). We have paid attention only to the so-called *essential characteristics* – '*characteristic* of a *concept* that is indispensable to understand that concept' (ibidem). As we will see, the distinctive characteristics of a concept are fundamental for creating concept systems and drafting definitions.

Throughout this work, we analyse the terms anchored in the double dimension of terminology (Costa 2013; Roche 2015), where we will reconcile iteratively, step by step, both the onomasiological and semasiological approaches. The onomasiological perspective makes us look at the concept designated by the term, identify it, isolate it, specify its characteristics to differentiate it from other concepts that belong to the same concept system. Finally, the concept is embedded in the concept system where it belongs. This approach is complemented by the traditional lexicographic methodology, which follows a semasiological path, in the sense that it begins from an existing corpus of specialised lexical units (the terms collected from the dictionary that will be referred to in the next section) to explore their semantic values. Following this mixed approach, only after the relations are well-established, will the lexicographer be able to propose a definition that must be validated by the domain expert. We have aimed to combine the conceptual perspective – i. e., knowledge organisation – with the linguistic perspective – focusing on the terms themselves by analysing the data extracted from the dictionary under study.

3. *Dicionário da Língua Portuguesa* as a case study

Our methodology has been applied to a scholarly dictionary of the Portuguese language now being developed by the Academia das Ciências de Lisboa – the *Dicionário da Língua Portuguesa* (DLP) (ACL 2021). This lexicographic work is a retro-digitised dictionary (Simões/Almeida/Salgado 2016) whose starting point was the *Dicionário da Língua Portuguesa Contemporânea* (DLPC) (ACL 2001), last published in 2001. Currently, it is being prepared under the Instituto de Lexicologia e Lexicografia da Língua Portuguesa's supervision in collaboration with researchers and invited collaborators. This project is supported by a small annual Community Support Fund Portuguese National Fund (Fundo de Apoio à Comunidade – FAC) through the Fundação para a Ciência e a Tecnologia (FCT). It will be the first academy Portuguese digital dictionary and it will soon be available online.

For illustrative purposes, we have selected some terms from the GEOLOGY domain, more specifically stratigraphical terms, taken from the DLPC. Stratigraphy is the branch of earth sciences that deals with stratified rocks. The OED defines it as 'the branch of geology concerned with the order and relative position of strata and their relationship to the geological timescale'. Saying 'the branch of' immediately conveys the idea of subordination to something. The OED definition allows us to say that `<Stratigraphy>` is a subordinate concept of `<Geology>`.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

The result of the application of the terminological methods gives rise to updated dictionary entries or senses for the DLP. Thus, the DLP has a double function: it will be both the corpus of analysis and the dictionary that will be improved with our methodological approach.

4. Terminological working methods for lexicographic work

Our methodological proposal has strictly lexicographic purposes and aims to employ terminological working methods to contribute to the treatment of specialised lexicographic content within general language dictionaries. The ultimate goal of our proposal is to offer strategies that can help lexicographers write accurate definitions. Meeting this need, we will address one of the most challenging tasks for any lexicographer – defining terms of subject fields they do not master.

The methodology we have followed assumes the completion of three essential stages: preparation, processing, and publishing. It is structured in ten phases to achieve the proposed objectives based on the theoretical assumptions mentioned before. Figure 2 presents the different phases that make up our methodology:

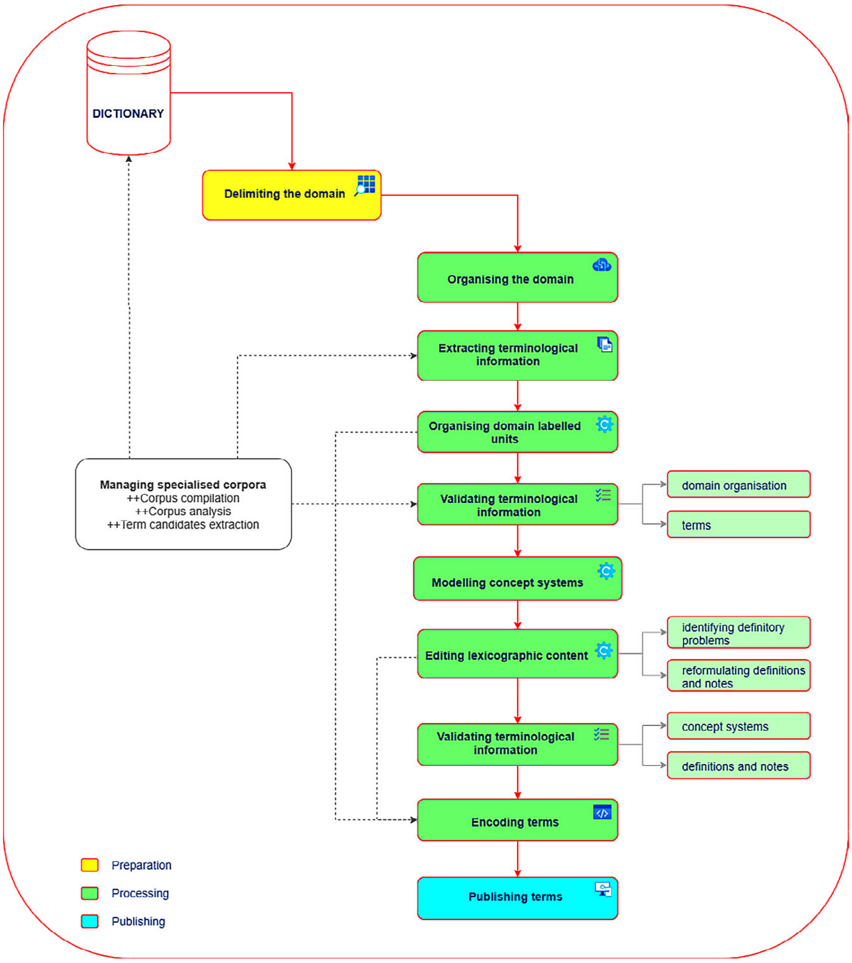

Fig. 2: Applying terminological methods when treating terms in general language dictionaries[2]

[2] For a detailed description of all the phases, see: Salgado (2021).

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

We can identify some tasks that have a purely linguistic nature, such as the analysis of terms as designations of concepts, and other extralinguistic tasks that have a conceptual nature (ISO 704 2009), e. g., delimiting and organising domains, identifying concepts and concept relations, and modelling concepts systems.

Throughout this paper, we focus on three stages: 1) Organising the domain; 2) Modelling concept systems; and 3) Editing lexicographic content.

4.1 Organising the domain

Getting to know the domain and subsequently organising it are two requisite activities for a swift and systematic identification of the concepts, which will result in a better description of the set of terms.

As mentioned above, domain knowledge-building in dictionaries is achieved by resorting to a set of domain labels. We have analysed DLCP’s labels and ended up suggesting the elimination of unnecessary or repetitive markings (Salgado/Costa/Tasovac 2021) and those distinctions that sometimes seem arbitrary because they are too narrow – both from a lexicographer’s point of view and that of a regular user of the dictionary.

Concerning the labels related to the GEOLOGY domain, we have found four domain labels related to the broader concept of `<EarthSciences>` in the DLPC (CRYSTALLOGRAPHY, GEOLOGY, MINERALOGY, and PALAEONTOLOGY). In the absence of an explanation of the domain labelling system in the DLPC front matter, we consulted some specialised literature and found these same labels/descriptors in other existing classification systems (e. g., Dewey Decimal Classification (DDC);[3] Universal Decimal Classification (UDC);[4] EuroVoc;[5] UNESCO Thesaurus)[6] – see Table 1.

DLPC	METALABEL	Dewey Decimal Classification (DDC)	Universal Decimal Classification (UDC)	EuroVoc	UNESCO Thesaurus
cristalografia	**crystallography**	540 Chemistry & allied sciences/548 Crystallography	54 Chemistry. Crystallography. Mineralogy/548/549 Mineralogical sciences. Crystallography. Mineralogy		36 Science/2.20 Physical sciences/Cristallography
geologia	**geology**	550 Earth sciences/551 Geology, hydrology & meteorology	55 Earth sciences. Geological sciences / 551 General geology. Meteorology. Climatology. Historical geology. Stratigraphy. Paleogeography	36 SCIENCE/3606 natural and applied sciences/NT1 geology	2 Science/2.35 Earth sciences/Geology
mineralogia	**mineralogy**	540 Chemistry & allied sciences/549 Mineralogy	54 Chemistry. Crystallography. Mineralogy/548/549 Mineralogical sciences. Crystallography. Mineralogy	36 SCIENCE/3606 natural and applied sciences/NT1 geology/NT2 mineralogy	2 Science/2.35 Earth sciences/Mineralogy
paleontologia	**palaeontology**	560 Paleontology; paleozoology	56 Paleontology		2 Science/2.35 Earth sciences/Paleontology

Table 1: Comparison of DLPC domain labels and existing classification systems (Salgado/Costa/Tasovac 2021)

3 https://www.oclc.org/en/dewey.html.

4 http://www.udcsummary.info/php/index.php.

5 https://eur-lex.europa.eu/browse/eurovoc.html.

6 http://vocabularies.unesco.org/browser/thesaurus.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Taking the DDC or UDC from Table 1 as an example, the domains of CRYSTALLOGRAPHY and MINERALOGY are indexed in the class that covers CHEMISTRY. The fact that these domains are associated with CHEMISTRY, not with GEOLOGY, is acceptable since much of the subject actually falls into the CHEMISTRY domain; however, one cannot disregard that the subject is also directly related to GEOLOGY. Thus, interdisciplinarity is always a central point for domain organisation, and we have to take this into consideration when organising specialised knowledge in dictionaries. In Table 1, we have a column entitled *metalabel*, a tag that identifies the equivalent English designation of the corresponding domain. Using a metalabel will be beneficial for any work on aligning multiple dictionaries and studying them in parallel. This metalabel will also play an important role in the domain hierarchy that we will propose later on for the benefit of annotation.

Comparing the different classification systems has allowed us to offer a proposal to represent domains associated with EARTH SCIENCES in general language dictionaries applied to the DLP (Fig. 3), which the expert we consulted validated.

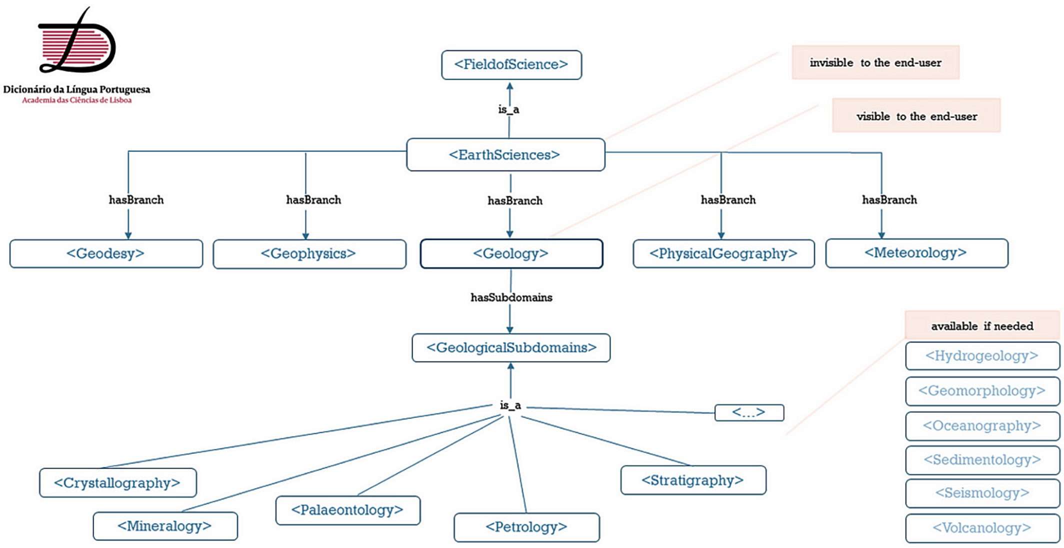

Fig. 3: Domain labels within the EARTH SCIENCES superdomain showing GEOLOGY as a domain and identifying its subdomains

This was the starting point to move from a non-hierarchical domain organisation to a hierarchical structure, which consequently increases the consistency of annotation and information retrieval. As Atkins/Rundell (2008) argued, instead of conceiving 'a totally flat non-hierarchical list of domains, it is more practicable to try to build a domain list with a certain hierarchical structure' (ibid., p. 184).

We have built hierarchical domain trees for lexicographic purposes. The hierarchy is as follows: superdomain, domain, subdomain (Salgado/Costa/Tasovac 2021). EARTH SCIENCES represent a superdomain followed by the domain GEOLOGY. In turn, GEOLOGY has various subdomains. Some domain labels will not be visible to the end-user since we consider labelling a lexicographic device for knowledge organisation. If the need to include other subdomains shall arise, they have already been foreseen. Our hierarchical domain trees can be made visible to give end-users the possibility of understanding the conceptual scope and how terms are interlinked, which is generally found isolated in general language dictionaries because they usually follow alphabetical order.

To encode hierarchical domain labels, we used a customised version of TEI for lexicographic datasets – TEI Lex-0 (Tasovac et al. 2018) – employing the mechanism for the definition of taxonomies already available in `<teiHeader>`. This is possible in both plain TEI and TEI Lex-0 but has not been documented until now as a solution for representing usage labels. With this approach, domain labels are documented in `<encodingDesc>` (encoding description). The domains established in the taxonomy are declared in `<classDecl>` (classification declarations). This element is used to group the source of the domain's taxonomy used by the header or elsewhere in the document. First, the `<taxonomy>` element identifies the structured taxonomy. The categories are documented in the `<category>` element. Category elements are described, each defining a single category within the given taxonomy. Then, child categories are defined by the contents of a nested `<catDesc>` (category description) element, which contains the designation of the domain in the identified language. A single category may contain more than one `<catDesc>` child, and can be described in different languages (`xml:lang`). As a result of this thought process, we can establish a multilingual hierarchy for the EARTH SCIENCES superdomain (Fig. 4):

```
<encodingDesc>
    <classDecl>
        <taxonomy xml:id="domain">
            <category xml:id="domain.earth_sciences">
                <catDesc xml:lang="en">Earth Sciences</catDesc>
                <catDesc xml:lang="pt">Ciências da Terra</catDesc>
                <catDesc xml:lang="es">Ciencias de la Tierra</catDesc>
                <catDesc xml:lang="fr">sciences de la Terre </catDesc>
                <category xml:id="domain.earth_sciences.geology">
                    <catDesc xml:lang="en">Geology</catDesc>
                    <catDesc xml:lang="pt">Geologia</catDesc>
                    <catDesc xml:lang="es">Geología</catDesc>
                    <catDesc xml:lang="fr">Geologie</catDesc>
                    <category xml:id="domain.earth_sciences.geology.mineralogy">
                        <catDesc xml:lang="en">Mineralogy</catDesc>
                        <catDesc xml:lang="pt">Mineralogia</catDesc>
                        <catDesc xml:lang="es">Mineralogía</catDesc>
                        <catDesc xml:lang="fr">Mineralogie</catDesc>
                    </category>
                </category>
            </category>
        </taxonomy>
    </classDecl>
</encodingDesc>
```

Fig. 4: Hierarchical domain label for EARTH SCIENCES domain labels

Flat usage labels are usually encoded as text values of the `<usg>` element. For the sake of human readability, one could deploy the same strategy and explicitly add the domain label as the content of the `<usg>` element even when the full label taxonomy is maintained in the `<teiHeader>`. This would be particularly useful when the labels used in a given dictionary are not consistent.

Having organised the domains, we can start working with concepts.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

4.2 Modelling concept systems

Understanding concepts and the terms that denote them accurately depends on understanding the concept relations that interlink concepts in a concept system[7]. Our references were the concept relations and the graphic representations in the UML (Unified Modelling Language) notation proposed by the ISO 704 (2009) standard through concept diagrams[8].

We have identified hierarchical relations – generic[9] and partitive[10] – and associative[11] relations.

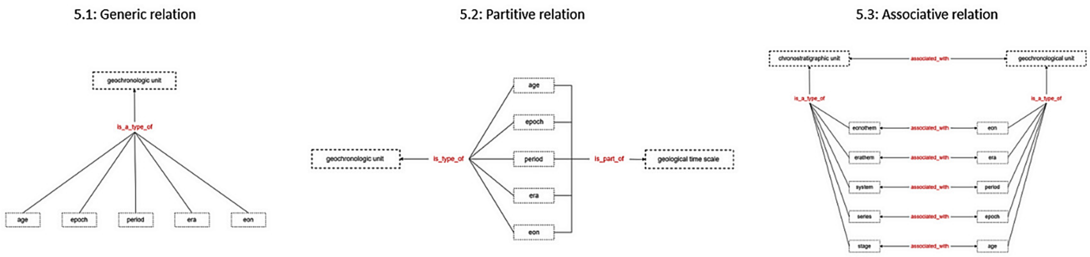

Fig. 5: Representation of conceptual relations using the concept of `<GeochronologicUnit>`

Here we have exemplified a generic concept relation (5.1) using the concept of `<GeochronologicUnit>` as a generic concept and `<Age>`, `<Epoch>`, `<Period>`, `<Era>`, and `<Eon>` as subordinate concepts. The specific concepts inherit a set of characteristics from their generic superordinate concept, i. e., the superordinate concept includes the subordinate concepts. The type of conceptual relation is made explicit using the marker `is_a_type_of`, which structures the generic/specific type relation. Regarding the semasiological approach, these markers also give us the possibility of detecting semantic relations, such as hypernym-hyponym relations. This exercise allowed us to detect that the superordinate concept `<GeochronologicUnit>` was not defined in the DLPC. Another argument required our attention: the subordination established between different concepts is not mirrored in the DLPC. These subordinate concepts constitute different entries in general language dictionaries, so one of the possible ways to represent the established semantic relations is to annotate them in TEI.

The primary means of conveying geological time information is through the Geological Time Scale and its units. Thus, all these units are part of the `<GeologicalTimeScale>` – a partitive relation (5.2). The conceptual relationship between the broader concept and its parts was made explicit through the conceptual marker `part_of`. Contrary to what was

[7] A concept system is understood as a 'set of *concepts* structured in one or more related *domains* according to the *concept relations* among its concepts' (ISO 1087 2019, p. 6).

[8] A concept diagram is a 'graphic representation of a *concept system*' (ISO 1087 2019, p. 7).

[9] A generic relation exists between two concepts when the intension of the subordinate concept includes the intension of the superordinate concept plus at least one additional delimiting characteristic (ISO 704 2009, p. 9).

[10] A partitive relation is said to exist when the superordinate concept represents a whole, while the subordinate concepts represent parts of that whole (ISO 704 2009, p. 13).

[11] An associative relation exists when a thematic connection can be established between concepts (ISO 704 2009, p. 17).

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

observed in generic relations, the principle of inheritance does not apply here, i. e., the concepts in a partitive relation do not inherit the characteristics of the superordinate concepts but do inherit their parts. The `<GeologicalTimeScale>` is a comprehensive concept, and all identified subordinate concepts – `<Age>`, `<Epoch>`, `<Period>`, `<Era>`, and `<Eon>` – represent parts of a whole, but they have distinctive characteristics concerning the related comprehensive concept.

To illustrate an associative concept relation (5.3), again we have used the concept of `<GeochronologicUnit>` in association with `<ChronostratigraphicUnit>`. We have a non-hierarchical relation: material–time, i. e., they have a semantic or pragmatic connection. If one wishes to allude to the time when these strata were deposited, then the concept of `<ChronostratigraphicUnit>` is replaced by that of `<GeochronologicUnit>`.

Once the conceptual relations are correctly identified, the lexicographer is able to start writing the definitions.

4.3 Editing lexicographic content

In this phase, senses are explained. For terminological purposes, a definition stabilises the relation between a concept and a term by the means of a linguistic expression. We distinguish the terminological definition (cf. De Bessé 1990; Rey 1995; Sager 2000; Temmerman 2000) from the lexicographic definition (Mel'čuk/Polguère 2018), which is generally suitable for general language dictionaries. Although both terminology and lexicography favour definition by intension, their purposes are different. The terminological definition attempts to state a concept designated by a term and characterise it in relation to other concepts within a concept system. In contrast, the lexicographic definition seeks to describe the signified meaning(s) of a lexical unit.

The terminological definition is related to the definition of the thing, as opposed to the lexicographic definition that relates to the usage of the word and is made by identifying the semantic features that characterise the meaning. The unit of meaning aimed at in the terminological definition is the concept. The difference between the terminological definition and the lexicographic definition, therefore, leads to different although not mutually exclusive approaches. In the context of general language dictionaries, the terminological definition has to be written for a non-expert audience.

ISO standards (ISO 704 2009; ISO 1087 2009) distinguish between intensional definition and extensional definition. The former consists of listing the immediate superordinate concept and delimiting the characteristics of the defined concept; the latter comprises listing its subordinate or partitive concepts. The definition by analysis or *genus-differentia* (Sager 1990) corresponds to ISO standards' intensional definition. Intensional definitions based on generic associations include the superordinate concept, followed by the delimiting characteristics within a concept system (e. g., `<Era>` among `<GeologicalTimeSpan>`). The superordinate concept's characteristics (that make up the intension) are assumed in the definition, which is the inheritance principle. Establishing conceptual relations facilitates the lexicographer's work and also enables the creation of a definitory model, e. g., `<GeochronologicUnit>` [superordinate concept] + `formed_during` [subordinate concepts].

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Existing definitions may have to be reformulated in cases where definitory problems may arise. On the other hand, the lexicographer can propose new definitions based on the previously established concept relations. For example, concerning the definition of "unidade geocronológica" [geochronologic unit], not included in the DLPC, we will suggest a definition considering the information retrieved from the following diagram (Fig. 6):

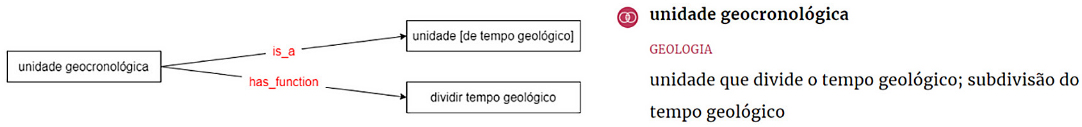

Fig. 6: Representation of the relation the conceptual markers `is_a` and `has_function` established from `<GeochronologicUnit>` and the unit "unidade geocronológica" defined in the DLP

As we can see in Figure 6, conceptual identifiers and linguistic markers may help lexicographers draft definitions. Focusing on the characteristics of a given concept is a fundamental step when defining it. The conceptual relation marker `is_a` establishes a hierarchical relation of subsumption. The conceptual marker `has_function` indicates the functionality of the unit. As we shall see later, we have assumed that these are instances of the so-called complex relationships (Sager 1990, pp. 34 f.), which are domain- and application- dependent. Thus, we have put forward the following definition for "unidade geocronológica" in the DLP: 'unidade que divide o tempo geológico; subdivisão do tempo geológico' [unit that divides geological time; geological time subdivision].

As lexicographers, we could not aim to work with all identified concepts. However, we consider it essential to analyse the relations among relevant concepts and organise them into concept systems, which will benefit the drafting of definitions. To illustrate this, Table 2 presents five different terms extracted from the DLPC and compares them with DLP's definitions that we have written after modelling the concept's microsystem. All of them define a type of `<GeochronologicUnit>`:

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

HEADWORD	DLPC (2001)	DLP (2021)
éon [eon]	*Geol.* longo período de tempo geológico que abarca duas ou mais eras	intervalo de tempo geológico (unidade geocronológica) durante o qual se formou um eonotema (unidade cronostratigráfica) Notas: 1) Na escala do tempo geológico, o éon é a categoria hierárquica mais elevada. 2) O éon integra várias eras.
era [era]	*Geol.* cada uma das grandes divisões do tempo geológico, cujos limites estão marcados por mudanças geológicas ou paleontológicas e que abrange vários períodos	intervalo de tempo geológico (unidade geocronológica) durante o qual se formou um eratema (unidade cronostratigráfica) Notas: 1) Na escala do tempo geológico, a era é hierarquicamente superior ao período e inferior ao éon. 2) A era integra vários períodos.
período [period]	—	intervalo de tempo geológico (unidade geocronológica) durante o qual se formou um sistema (unidade cronostratigráfica) Notas: 1) Na escala do tempo geológico, o período é hierarquicamente superior à época e inferior à era. 2) Na escala do tempo geológico, o período integra várias épocas.
época [epoch]	*Geol.* intervalo de tempo, nas divisões estratigráficas, que é relativo às formações de uma série ou conjunto de terrenos; subdivisão do período	intervalo de tempo geológico (unidade geocronológica) durante o qual se depositou uma série (unidade cronostratigráfica) Notas: 1) Na escala do tempo geológico, uma época é hierarquicamente superior à idade e inferior ao período. 2) Uma época integra várias idades.
idade [age]	—	intervalo de tempo geológico (unidade geocronológica) durante o qual se formou um andar (unidade cronostratigráfica) Notas: 1) A idade é a unidade básica da hierarquia do tempo geológico. 2) Quando necessário, a idade pode ser dividida em unidades geocronológicas de categoria inferior designadas por crono.

Table 2: Comparison of the definitions of “éon”, “era”, “período”, “época”, and “idade” in the DLPC (2001) and the DLP (2021)

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

For example, the lexicographic article “era” in the DLPC is ‘Cada uma das grandes divisões do tempo geológico, cujos limites estão marcados por mudanças geológicas ou paleontológicas e que abrange vários períodos’ [Each of the major divisions of geological time whose boundaries are marked by geological or palaeontological changes and which spans several periods]. This proposed definition lacks scientificity – it is too vague and even questionable as a formal statement. Comparing it to other Portuguese online dictionaries, we have found that “era” is defined in PRIBERAM (2021) as ‘Divisão da escala de tempo geológico, superior ao período e inferior ao éon’ [Division of geological time scale, higher than period and lower than eon] (PRIBERAM [emphasis added]). In INFOPÉDIA (2021), the lexicographic definition is ‘unidade de divisão de tempo geológico, hierarquicamente inferior ao éon e superior ao período, definida por critérios paleontológicos e litológicos’ [unit of geological time division, hierarchically lower than the eon and higher than the period, defined by palaeontological and lithological criteria] (INFOPÉDIA, 2021 [emphasis added]). On the contrary, and since we are modelling concept systems, we do not propose including this feature in the definition because the information given is not essential to define the given concept but may help to understand it. Instead, we recommend that additional information should be inserted as notes in the lexicographic article (in Table 2, see ‘Notas’ [Notes]).

Finally, if we observe the set of proposed definitions, the uniformity and systematisation in the treatment of terms are remarkable, highlighting the lack of systematisation in the previous edition. The analysis of the definitions according to the conceptual aspects is relevant in dictionaries even if the audience is not made up of experts.

5. Concluding remarks and further work

Following a terminologically-based approach improves the quality of the lexicographic product, both in terms of representation and organisation of knowledge and the description of terms themselves – the conceptual and linguistic dimensions. Combining these two different dimensions involves an iterative procedure. We should emphasise that we endorse the definition of the concept (ISO 1087 2019). In the DLP, we tested the creation of natural language definitions using concept systems. Focusing on the characteristics of a given concept is a fundamental step when defining it. We showed that conceptual identifiers and linguistic markers may help lexicographers draft definitions. As recommended by ISO 704 (2009), we conclude that intensional definitions are beneficial.

This work aims to facilitate the drafting of definitions, which, as we demonstrate, can be optimised and provided with greater scientific precision when we follow a terminological approach to the treatment of terms. The results obtained are immensely satisfactory, ensuring greater definition accuracy and quality. Instead of working a dictionary by classical alphabetical ordering (from A to Z), i. e., letter by letter, we found advantages in treating entries by sets of terms, first identifying the generic concept and describing its characteristics, and thus distinguishing it from other concepts.

By proposing hierarchical domain labels, we organise knowledge and establish higher and lower categories. The fact that we define a domain hierarchy does not mean that all proposed labels will be visible in the final product. This means that the lexicographer must structure the domains thoroughly and identify the terms according to the classification adopted. However, later on, the decision to make other domain categories visible to dictionary users must be weighed and considered taking into account the number of terms classi-

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

fied under that label and also looking at the set of tags and their statistics in the realm of an established superdomain.

Given TEI Lex-0 has a non-standard nature (yet), it can be changed to accommodate relevant dictionary structures. We intend to demonstrate that the results obtained are helpful for computational lexical encoding and can serve the purpose of natural language processing. We have shown that the currently recommended TEI Lex-0 practice of representing domain labels as flat values is not robust enough to deal with more complex, hierarchical domain structures. The proposal that we present here for encoding hierarchical domain labels can be used in any dictionary, including multilingual ones. We recognise, however, that it is only a starting point for what we consider to be a joint effort to standardise domain labels, and that we have only dealt with two domains with a sampling of examples in each. In the future, we are also interested in exploring the results in the field of ontology.

We will continue to invest in an effective trans-disciplinary approach that combines theories and methods of terminology and lexicography, and even other disciplines, placing best practice standards at the core of our research. Unquestionably, terminology, with its interdisciplinary nature, is integral to knowledge conceptualisation and organisation, which justifies our approach.

References

ACL [Academia das Ciências de Lisboa] (2001): Dicionário da Língua Portuguesa Contemporânea, 2 vols. Casteleiro, J. M. (Coord.). Lisbon.

ACL [Academia das Ciências de Lisboa] (2021): Dicionário da Língua Portuguesa. Salgado, A. (Coord.). Lisboa. [New digital edition under revision.]

Atkins, B. T. S./Rundell, M. (2008): The Oxford guide to practical lexicography. New York.

Béjoint, H. (1988): Scientific and technical words in general dictionaries. In: International Journal of Lexicography 1 (4), pp. 354–368.

Boulanger, J.-C. (2001): L'aménagement des marques d'usage technolectales dans les dictionnaires généraux bilingues, dans Les dictionnaires de langue française. Dictionnaire d'apprentissage, dictionnaires spécialisés de la langue, dictionnaires de spécialité. In: Bibliothèque de l'Institut de linguistique française. Études de lexicologie, lexicographie et dictionnairique, n. 4, pp. 247–271.

Cabré, M. T. (1994): Terminologie et dictionnaires. In: Meta 39 (4), pp. 589–597. doi:10.7202/002182ar.

Cabré, M. T. (1999): Terminology: theory, methods and applications. Amsterdam. doi:10.1075/tlrp.1.

Costa, R. (2013): Terminology and specialised lexicography: two complementary domains. In: Lexicographica 29 (1), pp. 29–42. Doi:10.1515/lexi-2013-0004.

Costa, R. (2021): Terminology in the digital age: the ontological turn: Part 2. TOTh Training School 2021, 1–2 June 2021, France, Université Savoie Mont Blanc. Bourget du Lac.

De Bessé, B. (1990): La définition terminologique. In: Chaurand, J./Mazière, F. (ed.): Actes du Colloque la Définition, organisé par le CELEX (Centre d´études du Lexique) de l'Université Paris-Nord (1988). Paris, pp. 252–261.

Estopà, R. B. (1998): El léxico especializado en los diccionarios de lengua general: las marcas temáticas. In: Revista de la Sociedad Española de Linguística 28 (2), pp. 359–387.

Faber, P. (2009): The cognitive shift in terminology and specialized translation. In: MonTi: Monografías de Traducción e Interpretación 1, pp. 107–134. doi:10.6035/MonTI.2009.1.5.

Felber, H. (1987): Manuel de terminologie. Paris.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Gaudin, F. (2007): Socioterminologie: une approche sociolinguistique de la terminologie. Bruxelles.

Guerra Salas, L./Gómez Sánchez, M. (2005): El léxico especializado en los diccionarios monolingües de ELE. In: Castillo Carballo, M. A./Cruz Moya, O./García Platero, J. M./Mora Gutiérrez, J. P. (eds.): Actas del XV Congreso de Asele. Las gramáticas y los diccionarios en la enseñanza del español como segunda lengua: Deseo y realidad. Sevilla, pp. 427–434. https://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/15/15_0425.pdf.

INFOPÉDIA (2021): Dicionário Infopédia da Língua Portuguesa. Porto Editora. https://www.infopedia.pt (last access: 24-03-2022).

ISO 1087 (2019): Terminology work – vocabulary – part 1: theory and application. Geneva.

ISO 704 (2009): Terminology work – Principles and methods. Geneva.

Mel'čuk, I./Polguère, A. (2018): Theory and practice of lexicographic definition. In: Journal of Cognitive Science 19 (4), pp. 417–470. doi:10.17791/jcs.2018.19.4.417.

Nomdedeu Rull, A. (2008): Hacia una reestructuración de la marca de 'deportes' en lexicografía. In: Azorín Fernández, D. et al. (Eds.), El diccionario como puente entre las lenguas y culturas del mundo. Actas del II Congreso Internacional de Lexicografía Hispánica. Alicante, pp. 764–770. https://dialnet.unirioja.es/servlet/articulo?codigo=5511595.

OED = Oxford English dictionary (2021). Oxford from https://www.oed.com (last access: 24-03-2022).

Paz Battaner, M. (1996): Terminología y diccionarios. In: Actes de la Jornada Panllatina de Terminologia. Barcelona, pp. 93–117.

PRIBERAM (2021): Dicionário Priberam da Língua Portuguesa. https://dicionario.priberam.org (last access: 24-03-2022).

Rey, A. (1985): La terminologie dans un dictionnaire général de la langue française: Le Grand Robert. In: TermNet News 14, pp. 5–7.

Rey, A. (1995): Essays on terminology. Amsterdam.

Roberts, R. P. (2004): Terms in general dictionaries. In: Bravo Gozalo, J. M. (ed.): A new spectrum of translation studies. Valladolid, pp. 121–140.

Roche, C. (2015): Ontological definition. In: Kockaert, H. J./Steurs, F. (eds.): Handbook of terminology. Vol. 1. Amsterdam, pp. 128–152.

Rondeau, G. (1984): Introduction à la terminologie. Montréal.

Sager, J. C. (1990): A practical course in terminology processing. Amsterdam.

Sager, J. C. (2000): Essays on definition. Amsterdam.

Salgado, A. (2021): Terminological methods in lexicography: conceptualising, Organising and encoding terms in general language dictionaries. Doctoral dissertation. Lisbon. https://run.unl.pt/handle/10362/137023.

Salgado, A./Costa, R./Tasovac, T. (2019): Improving the consistency of usage labelling in dictionaries with TEI Lex-0. In: Lexicography: Journal of ASIALEX 6 (2), pp. 133–156. doi:10.1007/s40607-019-00061-x.

Salgado, A./Costa, R./Tasovac, T. (2021): Mapping domain labels of dictionaries. In: Proceedings of XIX EURALEX International Congress: Lexicography for Inclusion. Alexandroupolis.

Simões, A./Almeida, J. J./Salgado, A. (2016): Building a dictionary using XML technology. In: Mernik, M./Leal, J. P./Oliveira, H. G. (Eds.), 5th Symposium on Languages, Applications and Technologies (SLATE'16) (14:1–14:8). Germany: Dagstuhl. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. doi:0.4230/OASIcs.SLATE.2016.0.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Tasovac, T./Romary, L./Bański, P./Bowers, J./Does, J./Depuydt, K./Erjavec, T./Geyken, A./Herold, A./Hildenbrandt, V./Khemakhem, M./Petrović, S./Salgado, A./Witt, A. (2018): TEI Lex-0: A baseline encoding for lexicographic data. Version 0.8.5. DARIAH Working Group on Lexical Resources. https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html.

Temmerman, R. (2000): Towards new ways of terminology description. The sociocognitive-approach. Amsterdam.

Tournier, J. (1992): Problèmes de terminologie en lexicologie anglaise et générale. Recherches en linguistique étrangère. XVI, pp. 215–226.

Wüster, E. (1979/1998): Introducción a la teoría general de la terminología y a la lexicografía terminológica. Barcelona: Institut Universitari De Lingüística Aplicada, Universitat Pompeu Fabra. [Einführung in die Allgemeine Terminologielehre und terminologische Lexikographie. Bonn 1979].

Contact information

Ana Salgado
NOVA CLUNL, Centro de Linguística da Universidade NOVA de Lisboa, Portugal/Academia das Ciências de Lisboa, Portugal
ana.salgado@fcsh.unl.pt

Rute Costa
NOVA CLUNL, Centro de Linguística da Universidade NOVA de Lisboa, Portugal
rute.costa@fcsh.unl.pt

Toma Tasovac
BCDH – Belgrade Center for Digital Humanities, Serbia
ttasovac@humanistika.org

Acknowledgements

This paper is supported by 1) the MORDigital – Digitalização do *Diccionario da Lingua Portugueza* de António de Morais Silva [PTDC/LLT-LIN/6841/2020] project financed by the Portuguese National Funding through the FCT – Fundação para a Ciência e Tecnologia and 2) the European Union's Horizon 2020 research and innovation programme under Grant agreement No 731015 (ELEXIS) (European Lexicographic Infrastructure).

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.